Systematic exploration of energy landscapes in stochastic simulators

Systematic exploration of energy landscapes in stochastic simulators

Eliodoro Chiavazzo*, Charles W. Gear†, Ioannis G. Kevrekidis†‡
*Politecnico di Torino/Energy Department, Turin, Italy
†Princeton University/Department of Chemical and Biological Engineering, Princeton, NJ, USA
‡Princeton University/Program in Applied and Computational Mathematics, Princeton, NJ, USA

***Abstract*— In this work, we present an approach for the exploration of low-dimensional effective potential landscapes. Making use of extrapolation in a low dimensional space of automatically learned variables (i.e. Diffusion Maps - DMAPs - variables) and machine learning schemes (e.g. Geometric Harmonics - GH) for lifting the new points into the ambient space, the described method enables to escape from local potential wells towards new minima. A simple three-dimensional stochastic differential equation system with a non-linear two-dimensional attractive manifold is considered for illustration purposes.**

I. Introduction

The dynamics of complex systems may be driven by the gradient of an energy function, so that major events in the behavior of the system can be as follows: i) trapping into deep energy wells; ii) transition between two minima passing through a saddle point owing to thermal agitation. A prototypical example of this kind of behavior is provided by proteins, whose configuration is often trapped into one of the local free-energy minima. Hence, when studying proteins, it is very important to identify all the relevant features of the free-energy landscape including the main energy wells, saddle points and minimum energy paths (MEPs) between nearby minima passing through the saddles. In general, molecular dynamics and Monte Carlo simulations are very inefficient tools for the exploration of the potential energy landscape owing to the fact that most of the computational time is spent in jiggling at the bottom of local minima. As a matter of facts, the energy barriers between minima mainly cause a trapping of the system configuration with transitions being pretty rare events. A plethora of methods have been suggested in the published literature for overcoming this issue [1], [2]. Here, following the equation free approach [3], we propose a new method capable to perform macroscopic tasks by properly initializing and using microscopic simulators.

II. The exploration approach

Starting form an arbitrary initial condition, a system of stochastic differential equations (SDEs) is let run for a sufficiently long time such that the solution trajectory gets into one of the potential wells. Upon removal of the early part, the remaining trajectory is composed by an initial point cloud sampling the potential well bottom.

Diffusion maps (DMAPs) can be used to extract a suitable low-dimensional parameterization of the latter point cloud (two-dimensional in the example of Fig. 1). An automatic

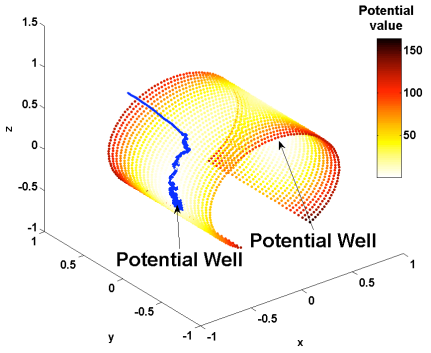

Fig. 1. (Color online) Two-dimensional attractive manifold. The color map refers to the value of a potential U and it clearly indicates that two wells are presents. An arbitrary solution trajectory will end in one of the two wells after a sufficiently long simulation time.

algorithm is needed for detecting the edge of the available point cloud. To this end, a few approaches have been suggested in the literature (see, e.g., the ball pivoting in [4]). For our purposes, we have formulated a simple enough procedure based on the notion of *maximum open angle* [5]. In general, the aforementioned procedure can be applied to both the physical (or ambient) and DMAPs space. For simplicity, in the following, we employ the algorithm to identify the cloud edge points directly in the DMAPs space (two-dimensional). Results (in DMAPs space) are shown on the right column of Fig. 2 by red circles. Upon detection of the entire edge (and provided that DMAPs do not provide an ill-posed parameterization [6]), boundary points can be safely extrapolated outwards in the DMAPs space. The green crosses on the right-hand side of Fig. 2 denote the extended boundary points at a fixed step of the suggested procedure. For obtaining a sufficiently uniform sampling of the extended boundary, the latter points (i.e. green crosses) are ordered and finely redistributed along the perimeter of the corresponding convex hull (using the readily available routine in Matlab: *convhull*). The reported one is just a possible implementation of the redistribution step, and more sophisticated algorithms may be also safely adopted.

The above operations provide us with a set of points in the two-dimensional DMAPs space and the final result is shown in the right-hand side of Fig. 2 by blue dots. However, the goal here is to re-initialize the detailed stochastic simulator in the

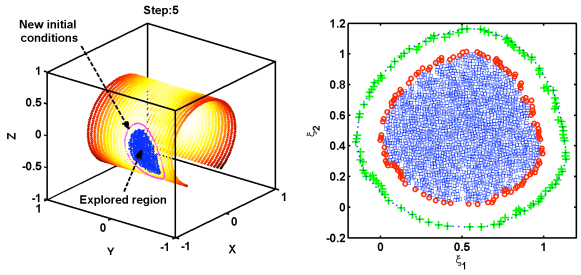

Fig. 2. An intermediate step of the suggested procedure for a three-dimensional stochastic system with a two-dimensional attracting manifold. On the left-hand side, the point cloud in physical (or ambient) space is reported whereas, on the right-hand side, the corresponding DMAPs parameterization of the explored region is shown (conveniently re-normalized so that the DMAPs variables: $0 < \xi_1 < 1$ and $0 < \xi_2 < 1$). Boundary points of the sample cloud (circles on the right-hand side) are automatically detected and outwardly extrapolated (crosses on the right-hand side) in the DMAPs space. A regular redistribution of the new extrapolated points is also performed by uniformly discretizing the perimeter of the convex hull containing the extrapolated points (see dots on the right-hand side). The redistributed points are finally lifted into the physical space, thus obtaining the new initial conditions to restart the process (see dots on left-hand side).

three-dimensional physical space, thus letting the available point cloud to invade larger portions of the phase-space towards possible unknown potential wells. Hence, relying upon the DMAPs parameterization, the redistributed extended boundary points can be *lifted* into the original full space by adopting a proper machine learning algorithm. Specifically, in our simulations, we utilized a local form of the geometric harmonics (GH) algorithm as discussed in [7], [8].

Thus, using the GH algorithm, the extended boundary is lifted in the three-dimensional space. Once the latter points are available, the process can restart by running the detailed simulator from the initial conditions indicated in the left-hand side of Fig. 2. Typically, after adding the new samples (i.e. trajectories starting from the *lifted* extended boundary initial conditions) to the previous point cloud, a *pruning* step is also performed, where only points sufficiently distant from their nearest neighbors are retained. This step ensures that the region of interest in the phase-space is sampled using a minimal number of points with fairly uniform point density.

III. Conclusions

Within the general framework of the equation free approach, we propose a new approach for the exploration of low-dimensional effective potential landscapes, as those occurring in stochastic simulators (e.g. molecular dynamics). By properly extrapolating along automatically learned coordinates (diffusion coordinates), it is possible to reach unexplored regions of the phase-space thus possibly discovering new energy minima (or minimum energy paths) of physical relevance. In general, the above procedure can be combined with a procedure for saddle point detection so that, when the latter event occurs, the system can be easily led into the new local minima and the process re-initialized from there.

Acknowledgment

E.C. wishes to acknowledge the support of the Italian Ministry of Research (FIRB grant RBFR10VZUG) and is grateful to G. Hummer for useful discussions and support.

References

[1] W. Zheng, M. Rohrdanz, C. Clementi, Rapid exploration of configuration space with diffusion map-directed-molecular dynamics, The Journal of Physical Chemistry B 117 (2013) 12769–12776.

[2] T. Frewen, G. Hummer, I. G. Kevrekidis, Exploration of effective potential landscapes using coarse reverse integration, The Journal of Chemical Physics 131 (2009) 134104.

[3] I. Kevrekidis, C. Gear, G. Hummer, Equation-free: The computer-aided analysis of complex multiscale systems, AIChE Journal 50 (2004) 1346–1355.

[4] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, G. Taubin, The ball-pivoting algorithm for surface reconstruction, IEEE Transactions on Visualization and Computer Graphics.

[5] S. Gumhold, X. Wang, R. MacLeod, Feature extraction from point clouds, In: International Meshing Roundtable, Sandia National Laboratories (2001) 293–305.

[6] C. Gear, Parameterization of non-linear manifolds, http://www.princeton.edu/ wgear/ 1–11.

[7] R. Coifman, S. Lafon, Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions, Applied and Computational Harmonic Analysis 21 (2006) 31–52.

[8] E. Chiavazzo, C. Gear, C. Dsilva, N. Rabin, Y. Kevrekidis, Reduced models in chemical kinetics via nonlinear data-mining, Processes 2 (2014) 112–140.